Sparsity Patterns

Michael Komodromos

April 11, 2022
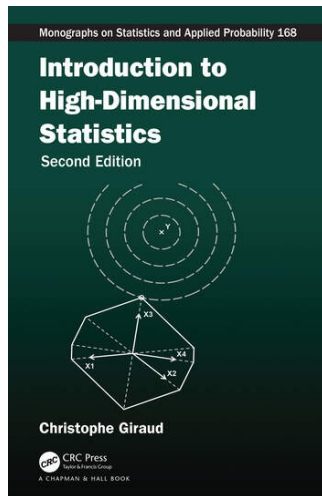
Outline

1 **Motivation**

2 **Sparsity Patterns**

3 **Examples**

4 **Questions?**

About this presentation

Based on the 2nd Chapter of "An
introduction to High-dimensional
Statistics" by Giraud (2021).

Can be found online for free

Motivation: Why should we care?

In many applications, particularly within this group, we work with high-dimensional data (more features than samples)

Using the data requires tools that take into account the patterns within it.

A commonly used tool is the LASSO, which assumes the data is sparse

The aim of this presentation is to introduce more of these tools

Motivation: A quick example

Consider the features produced by Texlab

There are groups of features describing different aspects of the image. For example,

`FOS_CV_LHH FOS_Imean_LHH FOS_Imedian_LHH FOS_Imode_LHH ...`

Group-sparse regression (which we'll look at later), is a way of incorporating this group information into the regression function.

Sparsity Patterns

Sparsity patterns: Statistical setting

Before we jump the different types of sparsity, we need to layout the statistical setting

Statistical setting

Formally, we'll be working in a regression setting where,

$$y_i = f(x_i) + \epsilon_i \tag{1}$$

which links our response $y \in \mathbb{R}$, to p variables stored in a p-dimensional real valued vector $x_i \in \mathbb{R}^p$.

Examples of $f(x)$

Example

Piecewise constant regression

$$f(x) = \sum_{j \in J} c_j \mathbb{I}(x \geq z_j) \tag{2}$$

Appears in the analysis of copy number variations

Example

Additive models

$$f(x) = \sum_{j \in J} f_j(x_j) \tag{3}$$

Main focus

However our main focus is going to be on linear regression models where
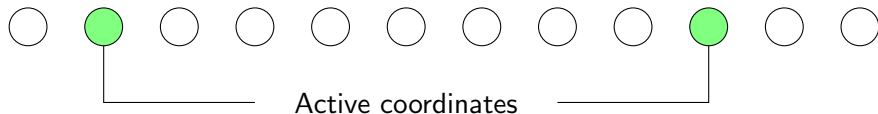
$$f(x) = \sum_{j \in J} \beta_j x_j \tag{4}$$

This is because all the previous examples (and many more) can be re-written in this way.

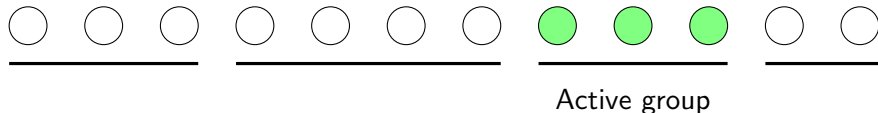# Sparsity patterns: Coordinate sparsity

Coordinate sparsity

Only a few coordinates of β are non-zero

Active coordinates

Sparsity patterns: Group sparsity

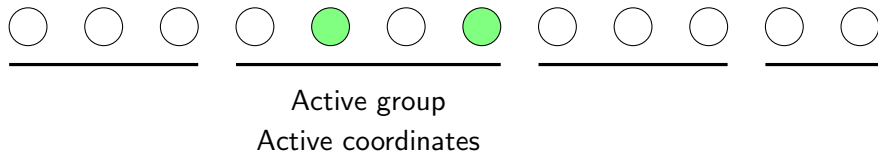## Group sparsity

The coordinates of β are clustered into groups, and only few of those groups are non-zero.

Active group

Sparsity patterns: Sparse-group sparsity

Sparse-group sparsity

The coordinates of β are clustered into groups, and only few of the coordinates within a group are non-zero.

Active group
Active coordinates

Sparsity patterns: in practise

The following R packages implement the different sparsity patterns and enforce sparsity via L_1 penalties

- Coordinate sparsity: `glmnet, seagull`

- Group sparsity: `gglasso, seagull`

- Sparse-Group sparsity: `ExclusiveLasso, seagull`

Examples

These examples are taken from

"A Sparse-group LASSO" by Simon et al. Sec 5.

A SPARSE-GROUP LASSO

NOAH SIMON, JEROME FRIEDMAN, TREVOR HASTIE, AND ROB TIBSHIRANI

ABSTRACT. For high dimensional supervised learning problems, often using problem specific assumptions can lead to greater accuracy. For problems with grouped covariates, which are believed to have sparse effects both on a group and within group level, we introduce a regularized model for linear regression with ℓ_1 and ℓ_2 penalties. We discuss the sparsity and other regularization properties of the optimal fit for this model, and show that it has the desired effect of group-wise and within group sparsity. We propose an algorithm to fit the model via accelerated generalized gradient descent, and extend this model and algorithm to convex loss functions. We also demonstrate the efficacy of our model and the efficiency of our algorithm on simulated data.

Keywords: penalize, regularize, regression, model, nesterov

Example: Breast cancer data

Dataset

- Gene expression values of $n = 60$ patients with estrogen positive breasts

- Patients were treated with tamoxifen for 5 years

- Classified according to whether cancer recurred

- After pre-processing $p \approx 12,000$ genes

- Genes are groups by cytogenetic position data (GSEA C1 data)

- 30 patients chosen at random used in the training set.
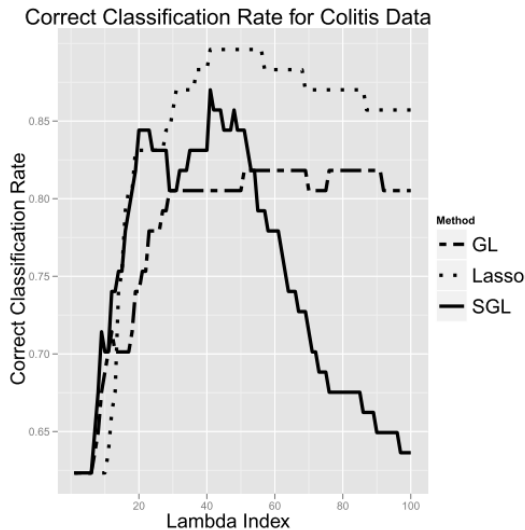
Example: Breast cancer data

Correct Classification Rate for Cancer Data
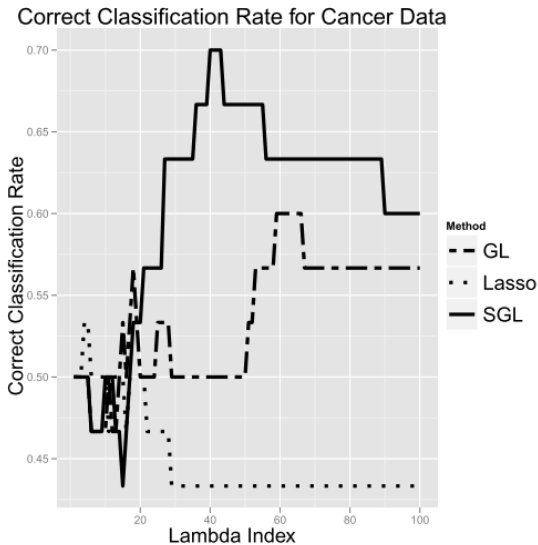
Example: Breast cancer data

Examining the peak classification accuracy for each method

Method	Classification Accuracy	Num. Features
Sparse group LASSO	70%	54 (11 groups)
Group LASSO	60%	74 (14 groups)
LASSO	53%	3

Example: Colitis data

- 127 patients

- 85 with colitis, 42 controls

- $p \approx 8,300$ after pre-processing

- grouped into gene-sets using cytogenetic information giving 277 disjoint groups.

- 50 observations used to fit models, 77 used to test.

Example: Colitis data

Correct Classification Rate for Colitis Data

Example: Colitis data

Examining the peak classification accuracy for each method

Method	Classification Accuracy	Num. Features
LASSO	90%	19
Sparse group LASSO	87%	43 (8 groups)
Group LASSO	84%	36 (7 groups)

In the second example the sparse group LASSO with the chosen gene sets did not perform as well as the LASSO - specialist information about the gene set may improve the results.

Take home message

Take home message

There are many different sparsity patterns for regression models.

Some may be more suited for solving problems than others.

When to use which? A rule of thumb

Method	Number of Groups	Size of groups
Coordinate sparse	Small	Large
Group sparse	Large	Small
Sparse group sparse	Large	Large

Reference I

[Gir21]   Christophe Giraud. *Introduction to High-Dimensional Statistics*. Chapman and Hall/CRC, Aug. 2021.

[Sim+13]   Noah Simon et al. "A sparse-group lasso". In: *Journal of Computational and Graphical Statistics* 22.2 (2013), pp. 231–245.